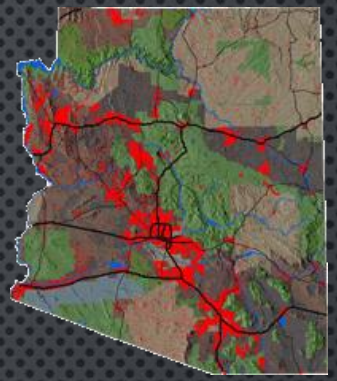


AZ-SMART

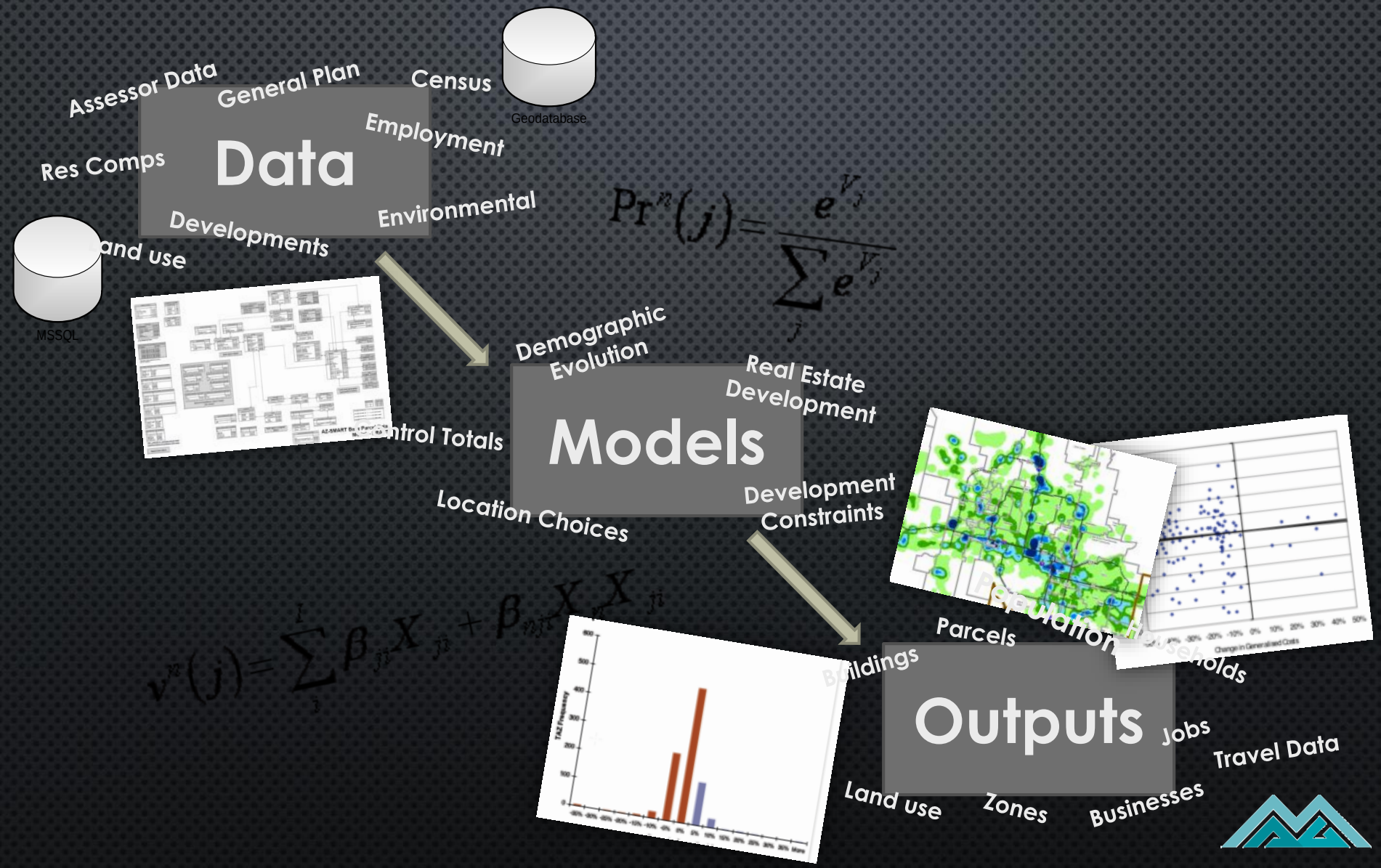
Arizona's Socioeconomic Modeling, Analysis, & Reporting Toolbox

AZ-SMART POPTAC WORKSHOP
FEBRUARY 23, 2016

AZ-SMART – OVERVIEW

- AZ-SMART: What and why?
- Input and Base Year Data
- Models: Simulation Sequence & Description
- 2016 Official Projections Review

AZ-SMART……..WHAT IS IT?



AZ-SMART……..WHAT IS IT REALLY?

- A framework for developing socio-economic projections
- Tools for data development and maintenance
  - Land use, general plan, inventories, etc.
  - Population and employment synthesis
- Tools for data analysis
- Tools for configuring, estimating, calibrating, and running simulation models
- Tools to assist in scenario building

AZ-SMART……WHY DO IT?

- We are required to produce population projections
- Travel demand model requirements
  - Travel models depend on socio-economic data
  - Socio-economic models are improved by travel model feedbacks
- We are getting asked questions that previous models were less equipped to answer, e.g.
  - Corridor studies
  - Activity-based travel models
- Policy analysis, what happens if:
  - Employment sector X is predicted to grow or decline substantially?
  - Several cities change land use policies to encourage density?

AZ-SMART Input and Base Year Data

DATA MODEL – TARGET DATA STRUCTURE

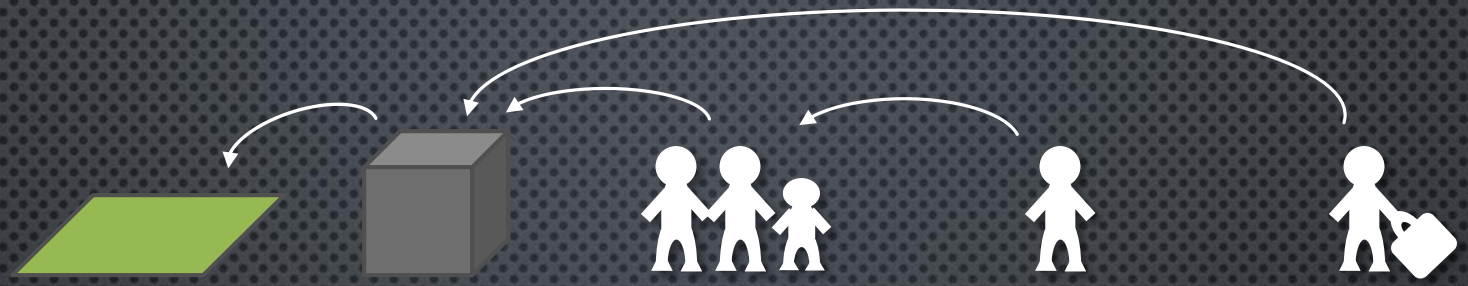
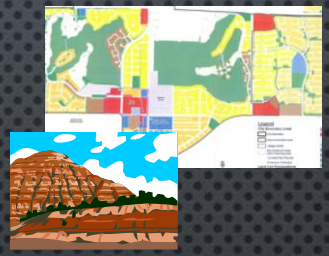


Table	Parcels	Buildings (built space)	Households	Persons	Jobs
Primary Key	zone_id	building_id	household_id	person_id	job_id
Foreign Key	taz_id, mpa_id, tract_id, etc.	zone_id	building_id	household_id	building_id
# Records	~1.42 million	~300 k	~1.44 million	~3.93 million	~1.79 million

Primary Input Databases
Physical and Built Environment

	Parcels / Zones	Built space	Constraints	Development Projects
				
Unique ID Field	Parcel id/ Zone id	Building id	Constraint id	Proposal id, Template id
Location Links	Zones, tract, cities, zip codes…	Parcel id / Zone id	Parcel id / Zone id	Parcel id / Zone id
Data Source	Census boundaries, Assessors parcels, etc.	Assessors data	City/ Town General Plan, Slopes, Flood plains, env. sensitive areas, etc.	Known development projects from City/Town

Primary Input Databases
Built space occupants / agents

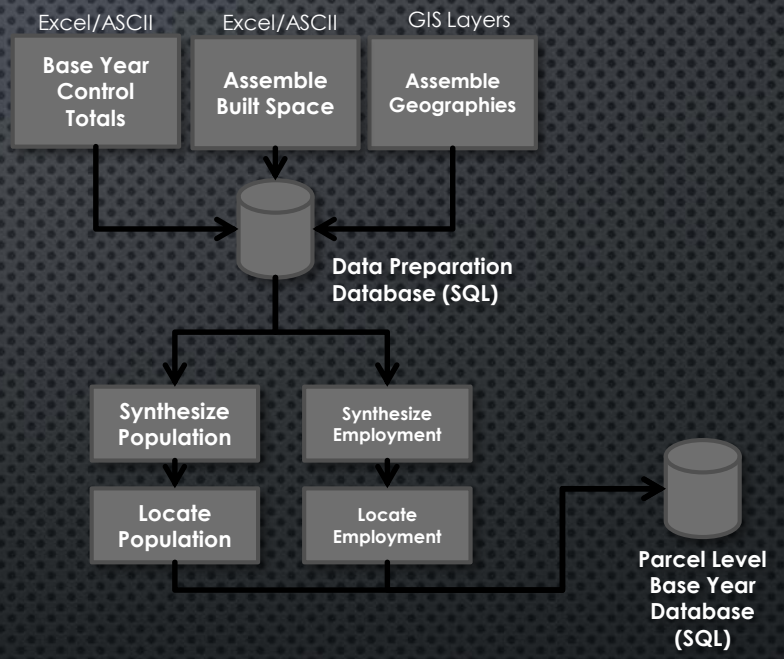
	Households	Persons	Businesses	Jobs
				
Unique ID Field	Household id	Person id	Business id	Job id
Location Links	Building id	Household id	Building id	Business id
Data Source	**Synthesized** from Decennial Census, American Community Survey (ACS), Public Use Microdata Samples (PUMS)		**Synthesized** from Employment data base from Quarterly Census of Employment and Wages (QCEW) / other proprietary sources	

PRIMARY INPUT DATABASES - REVIEWED

- Existing Land Use (EXLU) – updated and reviewed annually
- Developments (DEVS) – ongoing updates, reviewed annually
- General Plan (GP) – ongoing updates, reviewed annually
- MAG Employer Database (EmpDB) – data updated and reviewed annually
- Other Inventories (e.g. group quarters, schools, etc.) – updated and reviewed annually

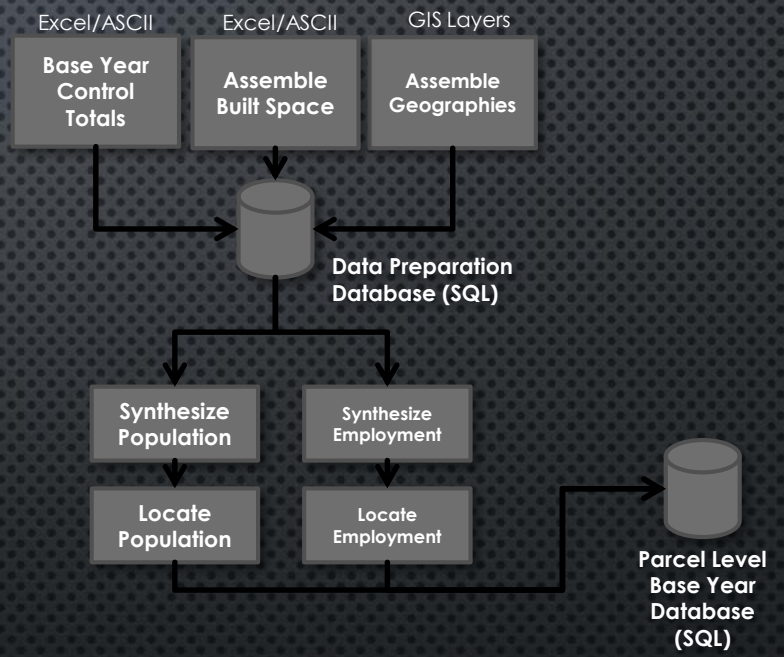
BASE DATA, ASSUMPTIONS, AND FACTORS

- DATABASE OF OBSERVED DATA REPRESENTING THE REGION
- DATABASE SERVES AS THE BASIS FOR FORECASTING, BUILD-OUT
- BASE YEAR IS 2014
  - NOTE: 2015 WILL BE A FORECAST YEAR DUE TO DATA AVAILABILITY, WILL BE CONTROLLED TO 2015 POPULATION ESTIMATE. ALL AVAILABLE 2015 DATA WILL BE USED.
- POPULATION CONTROL TOTALS
  - 2014 POPULATION ESTIMATES BY JURISDICTION
- EMPLOYMENT CONTROL TOTALS
  - 2014 TOTALS BASED ON 2014 Q1 AND Q2 QCEW
  - NEW 2014 TOTALS WILL BE USED FOR PROJECTIONS BASED ON FULL 2014 QCEW DATA



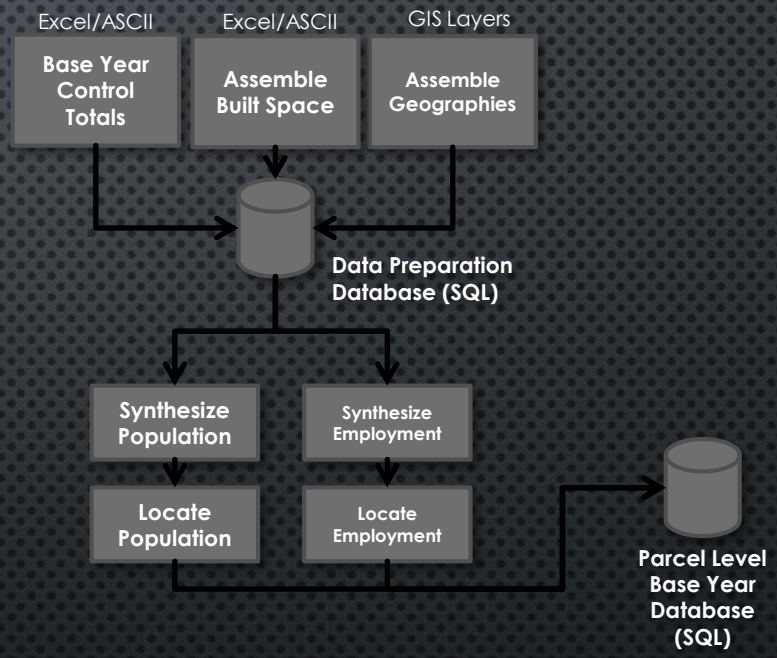
BASE DATA, ASSUMPTIONS, AND FACTORS

- LAND: ASSESSOR PARCELS, FILLED IN W/LAND USE DATASET
- BUILT SPACE: ASSESSOR IMPROVEMENTS, CITY AND OTHER PUBLIC DATA
  - BUILDING TYPES: 22 TYPES OF BUILT SPACE, AGGREGATED FROM ASSESSOR "MODEL TYPES" IN IMPROVEMENTS DATA
  - RESIDENTIAL COMPLETIONS CURRENT TO 2014Q2
- BUILDING OCCUPANTS:
  - HOUSEHOLDS
  - EMPLOYMENT
    - BUILDING SQFT / JOB, BY BUILDING TYPE
    - EMPLOYMENT BY INDUSTRIAL SECTORS: NAICS 2 DIGIT
    - WORK AT HOME EMPLOYMENT



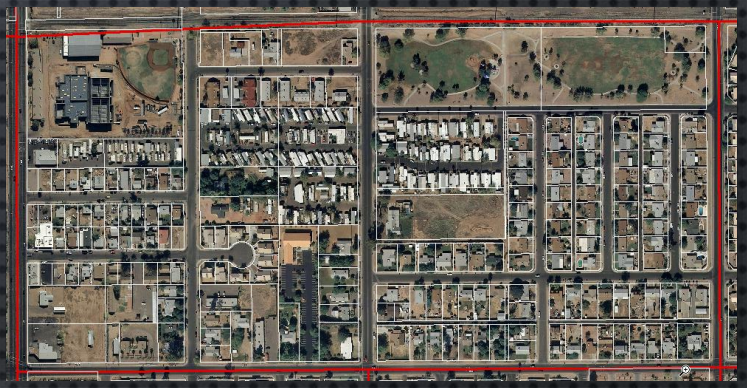
BASE DATABASE METHODOLOGIES

- HOUSEHOLD AND POPULATION SYNTHESIS:
  - BASED ON LATEST CENSUS ACS 5 YEAR SAMPLE AND ACS PUMS (2009-2013) RELEASED IN DEC. 2014
  - SYNTHESIZED TO "PSEUDO-BLOCK GROUPS," CONTROLLED POPULATION TO 2014 POPULATION ESTIMATES FROM ADOA
  - LOCATED HOUSEHOLDS TO DWELLING UNITS
- EMPLOYMENT SYNTHESIS:
  - 2014 MAG EMPLOYER DATABASE, PARCEL LEVEL
  - ALLOCATED EMPLOYEES TO BUILDING RECORDS PROPORTIONAL TO THE BUILT SPACE
  - SCALED EMPLOYMENT UP TO MATCH 2014 COUNTY CONTROL TOTAL (BY NAICS)
- SPECIAL POPULATIONS:
  - SEASONAL HOUSEHOLDS, ALLOCATED TO VACANT UNITS "HELD FOR SEASONAL USE" BY CENSUS
  - NON-SITE BASED (NSB) EMPLOYMENT, ALLOCATED TO BUILDINGS PROPORTIONALLY BY NSB TYPE AND BUILDING OCCUPANTS
  - CONSTRUCTION EMPLOYMENT, ALLOCATED TO DEVELOPING LAND USE CODES



POPULATION SYNTHESIS

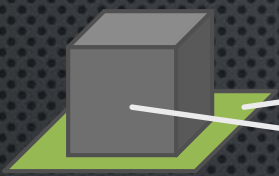
1. Given Census Block Group information (e.g. #hh by income categories & size, #pp by age and sex)
2. Generate frequency matrix (via IPF/IPU methodology) of household and person type constraints by block group
3. Draw from Census PUMS data detailed household and person records that match the constraints
4. Use detailed information from PUMS records (structure size, type, value, etc.) to match or "locate" households to Assessor parcels and built-space information with similar attributes



Total Households:		100
Households by Size:	1	30
	2	40
	3+	30
Households by Income:	Low	60
	High	40

Households		Household Income		Total
		Low	High	
Household Size	1	23.6	6.4	30
	2	15.2	24.8	40
	3+	21.3	8.7	30
Total		60	40	100

DATA MODEL – RESIDENTIAL EXAMPLE



Parcels

parcel_id	raz_id	mpa_id	county_id	area	freeway_within_qtr_mi	dist_to_phx_cbd	flood_plain
1671	204	26	1	174240	0	185042	0

Buildings

building_id	building_type_id	zone_id	land_area	residential_units	residential_units_capacity	non_residential_sqft	non_residential_sqft_capacity	sqft_per_unit
875	1	1671	174240	10	10	0	0	1964

Households

household_id	building_id	children	age_of_head	income	persons	workers
812943	875	1	60	122616	2	2
812960	875	1	30	79910	2	1
813534	875	0	39	44671	2	0
813556	875	0	38	37990	2	0
814097	875	0	71	74408	2	0
814118	875	0	65	76766	2	1
814258	875	0	31	93010	3	3
814287	875	0	67	129690	2	1
814434	875	0	25	0	1	0
814435	875	0	44	0	1	0

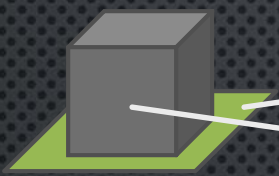
Persons

person_id	household_id	person_no	head_of_hh	age	sex	education	student_status	marriage_status
2170137	812943	1	1	60	1	11	1	1
2170138	812943	2	0	17	2	7	2	6
2170171	812960	1	1	30	1	8	1	1
2170172	812960	2	0	2	1	0	0	0
2171893	813534	1	1	39	1	5	1	1
2171894	813534	2	0	35	2	1	1	1
2171937	813556	1	1	38	1	9	1	1
2171938	813556	2	0	39	2	10	1	1
2173361	814097	1	1	71	1	9	1	1
2173362	814097	2	0	65	2	12	1	1
2173403	814118	1	1	65	2	14	1	1
2173404	814118	2	0	56	1	14	1	1
2173725	814258	1	1	31	1	10	1	6
2173726	814258	2	0	29	1	13	1	6
2173727	814258	3	0	28	1	13	1	6
2173857	814287	1	1	67	1	9	1	1
2173858	814287	2	0	59	2	9	1	1
2174154	814434	1	1	25	1	13	3	6
2174155	814435	1	1	44	1	11	1	6

Employment Synthesis

- Relies heavily upon the MAG employer database
  - Point level (parcel matched) database of employers
- Total employment in the employer database is less than the total employment present in the county
  - Employment must be scaled to match the county control
- Employer points match to built space / building records, many conflicts between datasets:
  - 1 employer point on a parcel with 1 building
  - 2+ employer points on a parcel with 1 building
  - 1+ employer point(s) on a parcel with 0 buildings
  - 2+ employer points on a parcel with 2+ buildings
  - Etc…

DATA MODEL – NON-RESIDENTIAL EXAMPLE



Parcels

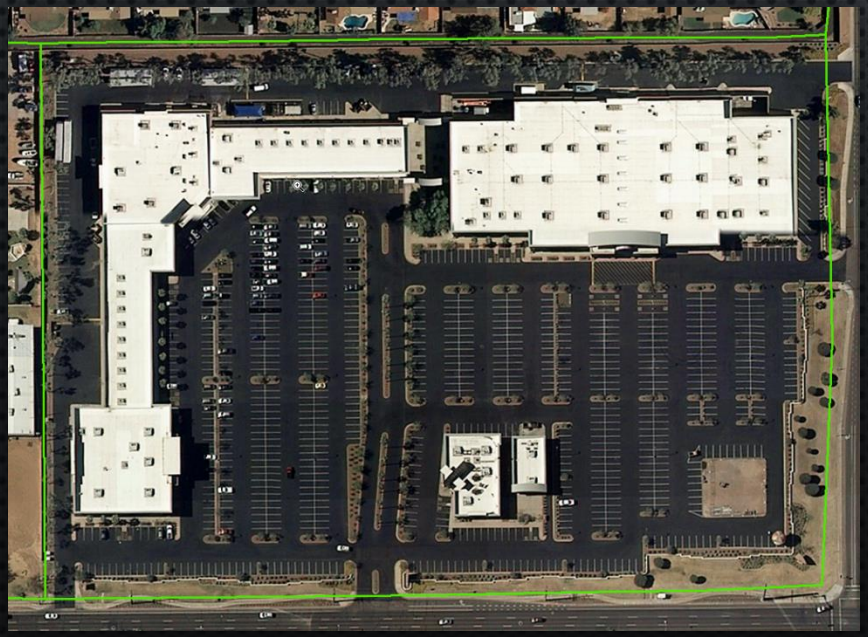
parcel_id	raz_id	mpa_id	county_id	area	freeway_within_qtr_mi	dist_to_phx_cbd	flood_plain
8479	265	27	1	849800	1	86950	0

Buildings

building_id	building_type_id	zone_id	land_area	residential_units	residential_units_capacity	non_residential_sqft	non_residential_sqft_capacity	sqft_per_unit
875	3	8479	849800	0	0	162500	162500	1

Jobs

job_id	building_id	sector_id	home_based_status
814097	875	3	0
814118	875	3	0
814258	875	3	0
814287	875	3	0
814434	875	3	0



building_type_id	building_sqft_per_job
1	150
2	150
3	800
4	400
5	350
6	500
7	500
8	750
9	1000
10	250

building_sqft_per_job

BUILDING TYPES

building_type_id	building_type_name	building_type_description
1	rsf	Single Family Detached Home
2	rmf	Multi Family Attached Home
3	mh	Mobile / Manufactured Home
4	retl	Retail
5	stor	MiniStorage
6	ware	Warehouse
7	ind	Industrial
8	off	Office
9	med	Medical
10	hot	Hotel
11	civic	Civic
12	edu	Education
13	gq	Group Quarters
14	pubf	Public - Federal
15	pubs	Public - State
16	publ	Public - Local
17	ag	Agriculture
18	tpt	Transportation
19	other imp	Other
20	os	Open Space

EMPLOYMENT SECTORS

sector_id	name	naics_code
1	Agriculture, Forestry, Fishing, Hunting	11
2	Mining, Quarrying, Oil, Gas	21
3	Utilities	22
4	Construction	23
5	Manufacturing	31-33
6	Wholesale Trade	42
7	Retail Trade	44-45
8	Transportation and Warehousing	48-49
9	Information	51
10	Finance and Insurance	52
11	Real Estate and Rental and Leasing	53
12	Professional, Scientific, Technical Svcs	54
13	Management of Companies and Enterprises	55
14	Admin., Support, Waste Mgmt, Remediation	56
15	Educational Services	61
16	Health Care and Social Assistance	62
17	Arts, Entertainment, and Recreation	71
18	Accommodation	721
19	Food Services and Drinking Places	722
20	Other Services (except Public Admin)	81
21	Public - Federal	92
22	Public - State	92
23	Public - Local	92

DATA MODEL – TARGET DATA STRUCTURE REVEIW

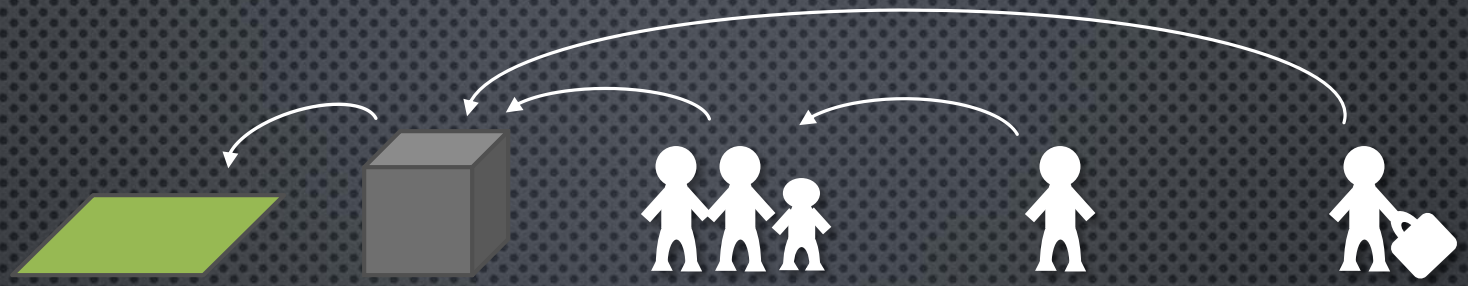


Table	Parcels	Buildings (built space)	Households	Persons	Jobs
Primary Key	zone_id	building_id	household_id	person_id	job_id
Foreign Key	taz_id, mpa_id, tract_id, etc.	zone_id	building_id	household_id	building_id
# Records	~1.42 million	~300 k	~1.44 million	~3.93 million	~1.79 million

Models: Simulation Sequence & Description

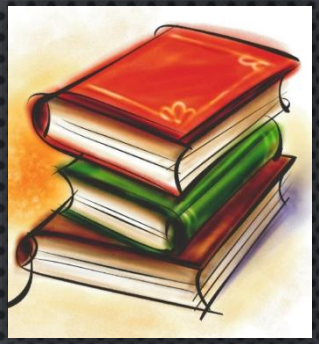
What is a model?

- A mathematical or statistical abstraction to help us understand and make predictions about real world systems, often consists of:
  - A set of simplifying assumptions
  - Initial conditions
  - A range of applicability
- Models can be used to:
  - Make predictions about the future
  - Understand what variables influence certain outcomes
- IMPORTANT: DETERMINISTIC VS PROBABILISTIC models
  - Deterministic: given a set of inputs, outputs are consistent over many model runs
  - Probabilistic (stochastic): has a random component, results vary some small amount over many model runs
  - AZ-SMART uses a combination of both types

AZ-SMART GENERAL MODEL TYPES

- **"Simple" model:**
  - Can be thought of as a simple calculation on tables (e.g. "SQL-like query")
- **Allocation models:**
  - Allocates a given total of 'X' to dataset 'Y' based on weights in 'Y' OR
  - Allocates A given number of 'X' agents to dataset 'Y' based on weights
- **"Rate Based Change" model:**
  - Randomly selects some percentage of some dataset based on pre-computed change rates
- **Regression model:**
  - Can be used to predict any continuous variable in a dataset during the simulation such as household income or real estate prices
- **Probability or Choice model:**
  - Discrete choice: binary (2 choice) or multinomial logit (2+ choices)

AZ-SMART MODELS: THEORETICAL BASES



- Random Utility Theory (McFadden)
  - Agents choose amongst alternatives based on relative utility
- Urban Economics/Bid Rent Theory (Alonso, Mills, Muth)
  - Explains land use as a function of willingness to pay for amenities and location trade-offs
- Hedonic Price Theory (Rosen)
  - Composite goods (housing) can be "disaggregated" into their attributes and priced
- Dynamic Market Equilibrium, Disequilibrium
  - Real-world markets rarely in equilibrium, imperfect information abounds
- Microsimulation (Orcutt)
  - Helps avoid aggregation bias and ecological fallacy
- GIS and Spatial Analysis (Tobler, Anselin)

Source: P. Waddell, 2010

AZ-SMART OVERALL SIMULATION SEQUENCE

Demographic Change Models

Employment Change Models

- Simulates demographic change in population and change in employment

Real Estate Development Models

- Predicts price, location, timing, and types of real estate development and re-development

Household Location Models

- Simulates migration and location choices of households

Employment Location Models

- Simulates location choices of regional employment

Demographic and Employment "Transition" Models

Demographic Change Models

Employment Change Models

- Simulates county level demographic change in population and change in employment
- Adds and removes households and jobs based on control totals developed a priori
- Randomly add or remove records until control totals are met, setting location to -1
- Results in new agents that need to be located to a building record

HOUSEHOLD AND EMPLOYMENT TRANSITION MODELS

Demographic Change Models

Employment Change Models

- Step 1: Calculate Simulation Population by bins in Control Totals Table and difference
- Step 2: Randomly duplicate or delete population records as necessary

Annual Population Control Totals Table

year	sex	age_min	age_max	Population Control Total	Simulation Population	Difference
2016	1	0	13	53684	52443	+ 1241
2016	1	14	17	52482	53451	- 969
2016	1	18	35	50472	48421	+ 2051
2016	1	36	55	49190	51216	- 2026
2016	1	56	75	48471	50994	- 2523
2016	1	76	-1	49621	48341	+ 1280
2016	2	0	13	48916	47154	+ 1762
2016	2	14	17	51751	50985	+ 766
2016	2	18	35	53478	52187	+ 1291
2016	2	36	55	52971	51924	+ 1047
2016	2	56	75	49168	51369	- 2201
2016	2	76	-1	47154	49271	- 2117

REAL ESTATE DEVELOPMENT MODELS

Generate Development Projects

Residential Unit Demand

Non-Residential "Unit" Demand

Choose Projects to Build

Generate Buildings for Chosen Projects

Real Estate Price Models

- In the first simulation year only: generate projects

- Parcels are grouped on their presence within a planned development or within a general plan area
- Large GP areas are subdivided by PLSS sections, parcels are sometimes split by various boundaries
- Only developable parcels are considered
- Parcels within a project are assumed to share the same future densities and land use compositions

REAL ESTATE DEVELOPMENT MODELS

- Generate Development Projects
- **Residential Unit Demand**
- **Non-Residential "Unit" Demand**
- Choose Projects to Build
- Generate Buildings for Chosen Projects
- Real Estate Price Models

- Estimate regional demand: determine how many units should be built county-wide for each development type

- Residential:
  - Calculate a total regional vacancy rate, compute # units to build based on a year specific target vacancy rate
  - Determine share of units going to single vs multi-family
- Non-residential:
  - Demand is expressed in terms of "job spaces" (building square footage / sqft per job)
  - Anticipate demand for building types based on new jobs' building types
  - Utilize target vacancy rate to compute # of job spaces to build

REAL ESTATE DEVELOPMENT MODELS

- Generate Development Projects
- **Residential Unit Demand**
- **Non-Residential "Unit" Demand**
- Choose Projects to Build
- Generate Buildings for Chosen Projects
- Real Estate Price Models

- Estimate regional demand: determine how many units should be built county-wide for each development type

- Residential:
  - Calculate a total regional vacancy rate, compute # units to build based on a year specific target vacancy rate
  - Determine share of units going to single vs multi-family
- Non-residential:
  - Demand is expressed in terms of "job spaces" (building square footage / sqft per job)
  - Anticipate demand for building types based on new jobs' building types
  - Utilize target vacancy rate to compute # of job spaces to build

Real Estate Development Models

- Generate Development Projects
- Residential Unit Demand
- Non-Residential "Unit" Demand
- **Choose Projects to Build**
- **Generate Buildings for Chosen Projects**
- Real Estate Price Models

- For each year:
  - Filter out complete or ineligible projects
  - Compute maximum # of units per project (based on size, type, # built)
  - Assign scores or weights to projects based on type and status
  - Choose top scoring projects by type

- Tier 1 projects: scheduled developments
  - have defined start and end years, guaranteed to build out at a linear rate
- Tier 2 projects: Active developments
  - Currently developing, subject to regional demand
  - More likely to be chosen to build, preference given to projects closer to completion and those with higher occupancy rates
- Tier 3: non-active projects
  - Not currently developing, subject to regional demand, less likely to be chosen

REAL ESTATE DEVELOPMENT MODELS

- Generate Development Projects
- Residential Unit Demand
- Non-Residential "Unit" Demand
- Choose Projects to Build
- Generate Buildings for Chosen Projects
- **Real Estate Price Models**

- Primary purpose: prices influence location decisions
- Model stratified into submodels by building type
  - Different variables influence SFR vs. Industrial prices
- Standard OLS estimated regression model with $/sqft as dependent variable
- Predicted prices then used as inputs to influence the behavior of other models (e.g. HLCM, ELCM)

PRICE = (B)HOME SQFT + (B)LOT SQFT + (B)WITHIN 1/4MI OF LANDFILL

Price	=	β	*	Home Sqft	+	β	*	Lot Sqft	+	β	*	0.25mi landfill?
Price	=	14.63	*	2365	+	2.34	*	43560	+	-5630	*	1
Price	=	34599.95			+	101930.4			+	-5630		
Price	=	$130,900.40										

AZ-SMART SIMULATION SEQUENCE

Household Location Models

Employment Location Models

- Simulates location choices of households
- Simulates location choices of regional employment

- **LCMs locate households and jobs in vacant residential and non-residential space**
  - Predicts a specific built-space record for a household or job in need of a location
- **Different types of households and employment have different variables that influence location decisions**
  - Household LCM stratified into models by 3 income classes
  - Employment LCM stratified into sectors (roughly corresponding to 2 digit NAICS)

AZ-SMART SIMULATION SEQUENCE – LCM VARIABLES

Household Location Models

Employment Location Models

- Accessibility variables using Travel Demand Model Skims:
  - <quantity> of <some variable> within <some travel time> <during some travel period> <on some mode>
  - Examples in current HLCMs / ELCMs:
    - Retail Employment within 20 minutes Peak SOV
    - Number of Jobs in Office buildings within 15 minutes Peak SOV
    - Resident Population within 15 minutes
    - Vacant Job spaces within 10 minutes
- Proximity variables:
  - Distance to nearest bus stop, light rail stop, park and ride, Dist to Phx CBD
- Other variables, examples:
  - HLCM: Interaction of household income with predicted DU price
  - DU sqft / persons in household
  - Zonal median income

LOCATION CHOICE MODEL EXAMPLE

1. Determine household records w/out a location
2. Identify vacant housing units
3. Randomly sample from vacant units for a reasonably sized choice set
4. Calculate utility and probability (based on household and housing unit characteristics)
5. Use random sampling to determine the choice

House	Utility	Probability
A	300	0.259
B	600	0.517
C	200	0.172
D	50	0.043
E	10	0.009

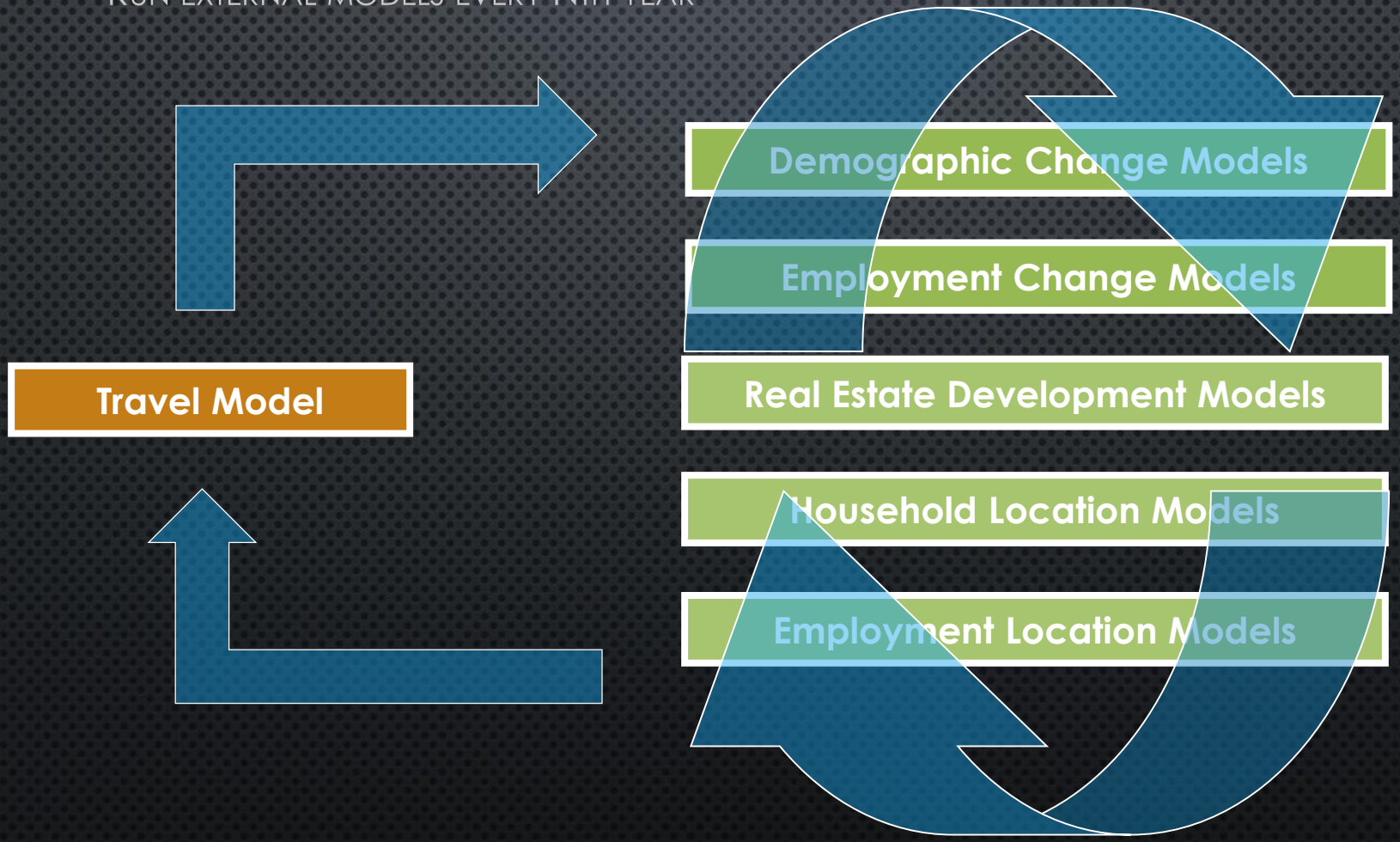
House	Ordered Probability	Cumulative Proability	Random Number
E	0.009	0.009	
D	0.043	0.052	
C	0.172	0.224	
A	0.259	0.483	0.480
B	0.517	1.000	

$$v^n(j) = \sum_i^I \beta_{ji} X_{ji} + \beta_{nji} X_n X_{ji}$$

$$\Pr^n(j) = \frac{e^{V_j}}{\sum_j e^{V_j}}$$

AZ-SMART SIMULATION SEQUENCE

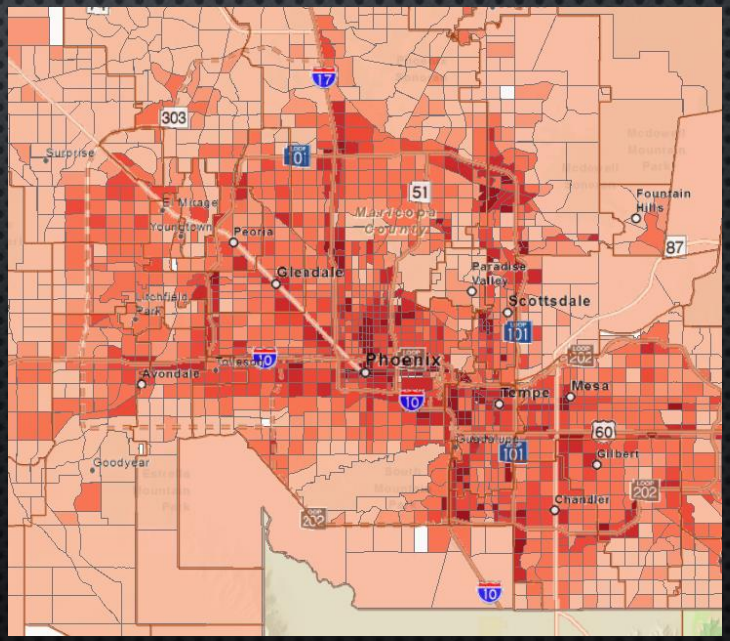
- Models repeat for every year
- Run external models every Nth year



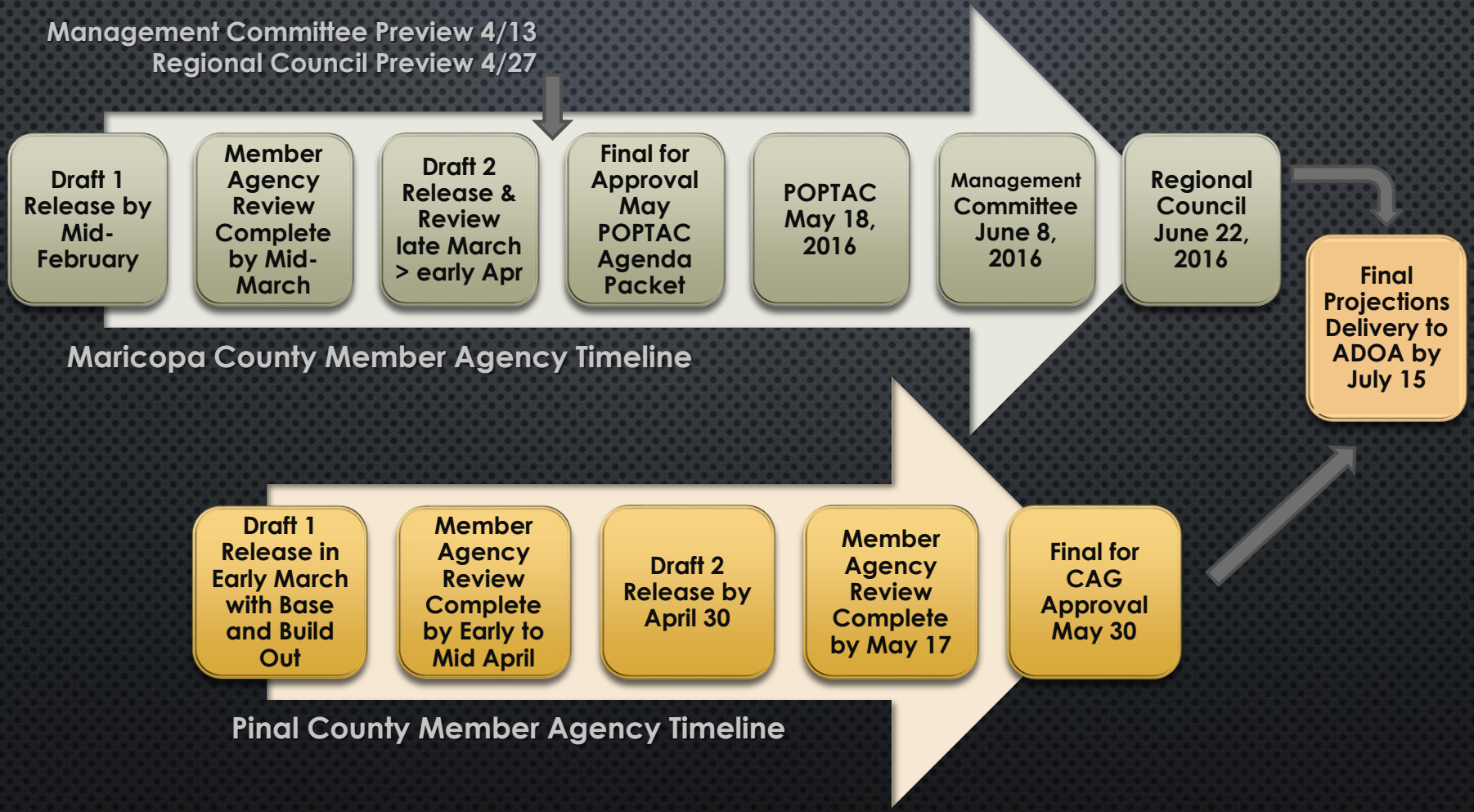
2016 Official Projections Review

Review Process

- Data inputs:
  - General Plan Land Use
  - Developments
  - Existing Land Use
  - Residential Completions
  - Hotels/Motels
  - Employer Database
  - Apartments/Mobile Homes/GQ
  - Job Centers
  - Municipal Planning Area Boundaries
- Member Agency Reviewers:
  - Planning
  - Economic Development
  - Transportation
  - Water Infrastructure
  - Others…
- Questions and Issues:
  - Development Trends
  - Public Services
  - Long-Term Infrastructure
  - Job Center Growth and Expansion

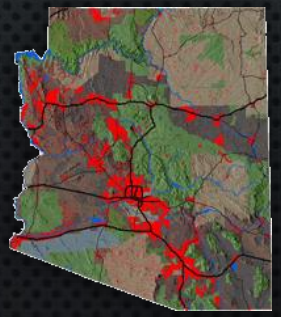


PROJECTIONS ACTIVITIES 2016



QUESTIONS?

PRIMARY TEAM:
HANYI LI
JESSE AYERS
SCOTT BRIDWELL
SCOTT WILKEN



JESSE AYERS
JAYERS@AZMAG.GOV

(602) 254-6300